灰色关联度分析在云南小春作物产量预报中的应用

顾本文, 吉文娟

(云南省农业气象与卫星遥感应用中心, 云南 昆明 650034)

摘要: 将灰色关联度分析法与相关分析法分别应用于云南小春作物产量预报因子的筛选, 并将用两种方法选出的因子用于2004年的产量预报。对比其预报结果, 灰色关联度分析法筛选出的气象因子, 更能体现其生物学意义; 可以通过计算灰色关联度来找到主要因子、剔除次要因子, 以抓住主要矛盾, 减少计算量。

关键词: 灰色关联度; 小春作物; 产量预报; 云南

中图分类号: S162.5 **文献标识码:** A **文章编号:** 1000-7601(2006)03-0045-04

对两个系统或两个因素之间关联性大小的量度, 称为关联度。灰色关联度分析是对于一个系统发展变化态势的定量描述和比较, 即关联分析是进行因素间时间序列的比较, 因此它的适用性广, 在综合评价、判别分析、类型区划等方面都有着广泛的应用, 并取得了很好的效果^[1]。

云南小春作物(即夏粮)主要指小麦、蚕豆、薯类和其他杂粮(如冬玉米等)。虽然其产量在全年粮食产量中所占比例不如秋粮, 但小春作物收成的好坏直接关系到农民是否能增收的问题。因为云南冬春季节气候资源优势明显^[2], 许多反季节粮蔬(如冬玉米、冬马铃薯等)在价格上有很强的竞争力, 所以准确预报小春产量, 能为政府部门调整产业结构、合理安排生产提供依据。

1 灰关联度原理与方法^[1]

1.1 数据变换

设有 m 个时间子序列和一个时间母序列, 分别表述为: $\{X_1^{(0)}(t)\}$ 、 $\{X_2^{(0)}(t)\}$ 、 $\dots$ 、 $\{X_m^{(0)}(t)\}$ 和 $\{X_0^{(0)}(t)\}$, 其中 $t = 1, 2, \dots, N$ 为序列的长度即数据个数, X 的上标(0)表示子序列和母序列均为未经任何处理的原始数据。由于子序列和母序列通常为不同种类的数据, 因此必须消除量纲后才能计算分析。目前, 原始数据变换有以下几种常用方法:

④ 均值化变换: 用序列的平均值去除对应序列中的各个原始数据, 所得到的新数据即为均值化序列。

⑤ 初值化变换: 用同一序列的第一个数据去除后面的各个原始数据, 所得到的新数据即为初值化数列。

⑥ 标准化变换: 将原始数据减去序列的平均值后再除以序列的标准差, 所得到的新数据即为标准

化序列。

1.2 计算关联系数

经数据变换的母序列记为 $\{X_0(t)\}$, 子序列记为 $\{X_i(t)\}$ 。在时刻 $t = k$ 时, 母序列 $\{X_0(k)\}$ 与子序列 $\{X_i(k)\}$ 的关联系数可由下式计算:

$$L_{0i}(k) = (\Delta_{\min} + \rho\Delta_{\max}) / (\Delta_{0i}(k) + \rho\Delta_{\max})$$

式中: $\Delta_{0i}(k)$ 表示 k 时刻母序列与子序列的绝对差, 即 $\Delta_{0i}(k) = |x_0(k) - x_i(k)|$ ($1 \leq i \leq m$); $\Delta_{\max}$ 和 $\Delta_{\min}$ 分别表示所有子序列与母序列比较中的各个时刻绝对差中的最大值与最小值; ρ 称为分辨系数, 其意义是削弱因最大绝对差数值太大引起的失真, 提高关联系数之间的差异显著性, $\rho \in (0, 1)$, 一般情况下可取 $0.1 \sim 0.5$ 。在本文的计算中取 0.5 。

1.3 计算关联度

两序列的关联度就是两个比较序列各个时刻的关联系数的平均值, 即:

$$r_{0i} = \frac{1}{N} \sum_{k=1}^N L_{0i}(k)$$

2 小春作物产量与气象要素的关联度分析

2.1 云南小春作物与气象条件

从生物学意义上讲, 小春作物的生长过程大致可以分为播种出苗期、生长中期、产量形成期、成熟收获期等。播种出苗期若雨水均匀适量, 则有利于发芽出苗和分蘖; 生长中期即拔节孕穗期, 是小麦一生中需水最多的时期, 充沛的降水能为穗多、穗大夺取高产打下良好基础; 产量形成期为孕穗灌浆期, 此时若遭遇低温冷害和干旱, 作物的空秕率增加; 成熟收获期若土壤水分严重不足, 则对小麦的生长及产量

收稿日期: 2005-10-06

作者简介: 顾本文(1967—), 女, 贵州省安顺市人, 硕士, 高级工程师, 主要研究方向为气候变化与预测, 气候资源区划。

形成非常不利^[3~4]。因此,本文选取小春作物整个生长期即 10 月中旬 ~ 翌年 4 月上旬各个旬段的雨量和 2 月中旬 ~ 3 月上旬各个旬段的 $\geq 0^{\circ}\text{C}$ 积温共 23 个因子与云南省小春单产作为一个系统来进行灰关联度分析。各气象因子依次用 $x_1, x_2, x_3, x_4, \dots, x_{22}, x_{23}$ 表示。

2.2 资料处理

将全省 1971 ~ 2003 年除西双版纳州、怒江州、迪庆州外的各气象站点的 23 个气象因子的序列资料的平均值作为子序列,云南省统计局统计 1971 ~ 2003 年的小春单产作为母序列。关联度的计算结果如表 1。

2.3 结果分析

从表 1 可看出,原始数据变换的方法不同,会影

响关联度的大小:初值变换的关联度值最大,均值变换的关联度值其次,标准变换的关联度值最小;三种方法的关联度值的排序也不同。分析前 10 个最优的因子可发现,尽管因原始资料的处理方法不同,入选的因子有差别,但仍有些因子是三者共有的:如 x_{20} (4 月下旬雨量)、 x_{21} (2 月中旬 $\geq 0^{\circ}\text{C}$ 积温)。这表明产量形成期 $\geq 0^{\circ}\text{C}$ 积温和成熟收获期雨量的确是影响小春产量的重要因子。

另外,按数理统计学的相关分析方法,计算了小春产量序列与 23 个气象因子序列的单相关系数(见表 1)。其统计量 $t = R \times [(n-2)/(1-R^2)]^{1/2}$ 遵从自由度为 $n-2$ 的 t 分布。在本文中,自由度 $n-2 = 28$,当显著水平 $\alpha = 0.05$ 时 $t_{\alpha} = 2.048$ 。

表 1 云南小春产量与各生育期气象因子的灰关联度
Table 1 The grey relation degree of spring crop yield forecast and growth period meteorological factors in Yunnan Province

气象因子 x Meteorological Factors	灰关联系数 r Grey relation degree			单相关系数 R Simple correlation coefficient	统计量 t Statistic
	均值化变换 Equalization transformations	初值化变换 Initial value transformations	标准化变换 Standardization transformations		
x_1	0.91419	0.96755	0.74276	0.27824	1.53281
x_2	0.88506	0.98239	0.72627	0.02732	0.14464
x_3	0.87096	0.97259	0.74153	0.06525	0.34602
x_4	0.85376	0.98008	0.68339	-0.26021	-1.42602
x_5	0.87077	0.94413	0.73338	-0.01695	-0.08972
x_6	0.87258	0.97827	0.75873	0.26568	1.45824
x_7	0.85775	0.98093	0.73699	0.09307	0.49463
x_8	0.84906	0.97945	0.75621	0.09170	0.48731
x_9	0.86345	0.98157	0.74821	0.32743	1.83370
x_{10}	0.86608	0.93755	0.75448	0.15800	0.84668
x_{11}	0.85610	0.97625	0.77084	0.32388	1.81148
x_{12}	0.85936	0.98174	0.69931	-0.07500	-0.39798
x_{13}	0.86412	0.91405	0.71726	0.14064	0.75167
x_{14}	0.84505	0.91992	0.78305	0.41446	2.40985
x_{15}	0.85838	0.82044	0.73201	-0.12032	-0.64131
x_{16}	0.87581	0.97402	0.75015	0.16113	0.86392
x_{17}	0.88140	0.98374	0.73911	0.29952	1.66115
x_{18}	0.87672	0.98363	0.69691	-0.23528	-1.28096
x_{19}	0.88999	0.97092	0.67900	-0.33397	-1.87487
x_{20}	0.91966	0.98309	0.74832	0.01879	0.09943
x_{21}	0.96431	0.99156	0.77051	0.25312	1.38447
x_{22}	0.95268	0.99218	0.69987	-0.11202	-0.59650
x_{23}	0.95775	0.99248	0.71110	0.12597	0.67194

从表 1 可看出 23 个气象因子中除第 14 个因子(2 月中旬雨量)外,其余因子与小春产量序列的 t

值均小于 t_{α} ,因此从统计学相关分析的角度看,认为绝大多数气象因子与小春产量序列关系不显著。但

从生物学意义上分析,此结论是不符合实际的。比如从抽穗到成熟恰恰是小麦一生中的最大需水期^[5],因此对云南这样一个山区面积占全省总面积 94%的省份来说,3 月上旬~4 月下旬小麦成熟收获期自然降水的多寡是至关重要的。可见,在分析云南小春作物产量与气象因子的相关性上,灰关联度分析比统计学相关分析能更好地体现两者间的生物学意义。由此我们得到启发:衡量两个因子之间的关系,既可以从数理统计的相关分析出发,计算其相关系数;也可以从灰色系统理论的关联度分析出发,计算其关联度;从不同方面多角度分析,才能判断出两个量之间真正的关系。

3 2004 年云南小春产量预测

从灰色系统理论出发,将均值变换、初值变换、标准变换挑选出的前 10 个最优的气象因子(见表 2)和从数理统计理论出发,前 10 位相关系数所对应的气象因子也作为预报因子,用“农业气象产量预报业务系统(版本 V2.0.1,中国气象科学研究院农业气象和遥感应用研究中心 1999—2001)”对 2004 年的小春产量进行了预测,其对应的预报值即社会产

量分别由 Y_1 、 Y_2 、 Y_3 、 Y_4 表示,结果见表 3。2003、2004 年云南小春实际产量分别为 $2\ 047.72\ \text{kg}/\text{hm}^2$ 和 $2\ 008.2\ \text{kg}/\text{hm}^2$ 。

“农业气象产量预报业务系统(版本 V2.0.1)”主要用于开展作物产量业务预报工作,作为一种基本的业务平台已经在全国气象部门推广应用。系统中可以使用多种预测因子,如本地区的地面气象因子气温、光照和降水等,利用这些预测因子制作不同的产量预报模型,通过对趋势产量和气象产量的模拟,从而得到实际的社会产量(社会产量=趋势产量+气象产量)。

表 2 前 10 个最优因子的灰关联度排序

Table 2 The collate of the optimal factors in grey relation degree

灰关联系数 Grey relation degree	关联度排序 Collate
均值化变换 Equalization transformations	$r_{21} > r_{23} > r_{22} > r_{20} > r_1 > r_{19} > r_2 > r_{17} > r_{18} > r_{16}$
初值化变换 Initial value transformations	$r_{23} > r_{22} > r_{21} > r_{17} > r_{18} > r_{20} > r_2 > r_{12} > r_9 > r_7$
标准化变换 Standardization transformations	$r_{14} > r_{11} > r_{21} > r_6 > r_8 > r_{10} > r_{16} > r_{20} > r_9 > r_1$

表 3 3 组最优预报因子的预报结果

Table 3 The result in forecasting of three groups optimal factors

方法 Methods	均值变换 Equalization transformations	初值变换 Initial value transformations	标准变换 Standardization transformations	单相关系数 Simple correlation coefficient
预报因子 Forecast factors	x_{21} 、 x_{23} 、 x_{22} 、 x_1 、 x_{20} 、 x_{19} 、 x_2 、 x_{18} 、 x_{17} 、 x_{16}	x_{23} 、 x_{22} 、 x_{21} 、 x_1 、 x_{13} 、 x_{15} 、 x_{19} 、 x_{14} 、 x_{16} 、 x_{10}	x_{14} 、 x_{11} 、 x_{21} 、 x_6 、 x_8 、 x_{10} 、 x_{16} 、 x_{20} 、 x_9 、 x_1	x_{14} 、 x_9 、 x_{11} 、 x_{17} 、 x_1 、 x_6 、 x_{21} 、 x_{16} 、 x_{10} 、 x_{13}
预报方法 Forecast methods	逐步回归 Stepwise regression	逐步回归 Stepwise regression	逐步回归 Stepwise regression	逐步回归 Stepwise regression
因子处理方法 Factors treatment methods	线性组合、距平值、 平方、对数 Linear combination、 departure、square、 logarithm	线性组合、距平值、 平方、对数 Linear combination、 departure、square、 logarithm	线性组合、距平值、 平方、对数 Linear combination、 departure、square、 logarithm	线性组合、距平值、 平方、对数 Linear combination、 departure、square、 logarithm
复相关系数 Multiple correlation coefficient	0.712	0.612	0.530	0.617
历史拟合率 Historical fitting coefficient	93.66%	93.80%	92.85%	93.75%
2004 年的预测值 The predicted value of 2004	2001.27 kg/hm^2 (-2.27%)	1997.27 kg/hm^2 (-2.47%)	2089.13 kg/hm^2 (+2.02%)	2072.29 kg/hm^2 (+1.20%)
预测模型 Forecasting model	$Y_1 = -203.181 + 0.02533 X_2^2 - 0.28646 X_{19}^2 + 60.45820 \ln X_{19} + 5.43175 X_{d16}$ $Y_2 = -123.977 + 0.02274 X_2^2 - 17.22045 X_9 + 0.86422 X_9^2 + 55.14155 \ln X_9$ $Y_3 = -83.273 + 2.38655 X_{10} + 22.33639 \ln X_{10} + 2.46470 X_{14} + 0.18624 X_{14}^2$ $Y_4 = -91.164 + 0.21845 X_{14}^2 + 0.22978 X_9^2 + 0.33401 X_{13}^2 + 25.89788 \ln X_9$			

注:括号内的数值为 2004 年的预测值与 2003 年实际值的趋势比较。预测模型中 X_{d16} 为因子的距平值。

Note: The numerical value in bracket is compare of predicted value in 2004 and actual value in 2003. X_{d16} is factor's departure in forecasting models.

本文中 4 个气象因子产量预报模型,对预报因子均进行了线性组合、距平值、平方和对数(以 e 为底)函数的处理;趋势产量模拟采用线性滑动平均和调和权重方法,滑动步长取 15;气象产量用逐步回归,单相关水平值取 0.25,滑动相关方差取 0.25,因子引入水平 2.5。

从表 3 可看出,在建模方法相同的情况下,用灰色系统和数理统计法选出的气象因子所做的预报结果是不相同的;在复相关系数方面,灰色系统(如均值变换)高于数理统计;但历史拟合率却是数理统计的略高于灰色系统的;在预报趋势上,灰色系统(如均值变换、初值变换)的明显好于数理统计的;在预报精度上,灰色系统的均值变换预报结果最接近实际产量。这说明,在一个系统中,灰关联度分析能把握主要矛盾,找出影响最大的因子;用灰关联度分析剔除影响不大或不太重要的因子后,对系统不产生本质的影响。

4 结论

1) 灰关联度分析比相关分析更能抓住主要矛

盾。实际应用中,若相关系数不明确,不能说明系统与因子、或因子与因子之间的关系时,可进一步通过计算灰关联度来分析问题。

2) 原始资料的处理方法不同,灰关联度的大小和排序受影响。

3) 用灰关联度分析剔除影响不大或不太重要的因子后,对系统不产生本质的影响。在实际应用中,若因子太多,计算量太大,可先通过灰关联度分析,减少因子个数(即降维),再进行下一步的研究。

参考文献:

- [1] 唐启义,冯明光.实用统计分析及其 DPS 数据处理系统[M].北京:科学出版社,2002.
- [2] 王宇.云南省农业气候资源及区划[M].北京:气象出版社,1990.31-32.
- [3] 邹丽云.云南小春生产的主要农业气象问题[J].云南气象,1992,(1):31-33.
- [4] 邹丽云.云南小春产量农业气象预报业务技术探讨[J].云南气象,1990,(3):22-24.
- [5] 刘汉中.普通农业气象学[M].北京:北京农业大学出版社,1990.195-196.

Application of grey relational analysis in yield forecast of spring crop in Yunnan Province

GU Ben-wen, JI Wen-juan

(Yunnan Center of Agrometeorology and Remote Sensing Application, Kunming, Yunnan 650034, China)

Abstract: We filtrate forecast factors of spring crop yield in Yunnan Province using grey relational analysis and correlation analysis. The filtrated factors are used for yield forecast of 2004. Through comparative analysis we can educe conclusion; using the method of grey relational analysis, the meteorological factors can incarnate biological meaning more profitably; we can find out major factors in a system, eliminate minor factors, and grasp the principal contradiction to decrease computational complexity.

Keywords: grey relation; spring crop yield; forecast